

Semicustom Design Methodology of Power Gated Circuits for Low Leakage Applications

Hyung-Ock Kim, *Student Member, IEEE*, and Youngsoo Shin, *Senior Member, IEEE*

Abstract—The application of power gating to cell-based semicustom design typically calls for customized cell libraries, which incurs substantial engineering efforts. In this brief, a semicustom design methodology for power gated circuits that allows unmodified conventional standard-cell elements is proposed. In particular, a new power network architecture is proposed for cell-based power gating circuits. The impact of body bias on current switch design and the layout method of current switch for flexible placement are investigated. The circuit elements that supplement cell-based power gating design are then discussed, including output interface circuits and state retention flip-flops. The proposed methodology is applied to ISCAS benchmark circuits and to a commercial Viterbi decoder with 0.18- μm CMOS technology.

Index Terms—Leakage, low power, power gating, semicustom, standard cell.

I. INTRODUCTION

SUBTHRESHOLD leakage current grows exponentially with every process generation due to the scaling down of the threshold voltage. Many circuit level approaches have been proposed. These include input vector control, power gating, dynamic voltage scaling, and body biasing. Power gating [1]–[3] refers to gating, or cutting off, a circuit from its power supply rails during standby mode. It has been widely used in the semiconductor industry to reduce subthreshold leakage as well as other leakage components.

Power gating is realized by placing a current switch, called a header, in series with a logic block, as shown in Fig. 1. A footer, which is a nMOS switch placed between the logic block and V_{ss} , can also be used. When the power management unit (PMU) detects a sufficiently long period of idle time, it turns off the header to disconnect the logic block from the power rail V_{dd} . When it subsequently detects that the logic block is required, the PMU turns on the header again so that the logic block is reconnected to the power rails. The rail between the logic block and the header, denoted by V_{ddv} in Fig. 1, serves as a virtual power rail for the logic block, which usually employs a low threshold voltage (V_t) to sustain its performance. The header, however, can have either a low V_t or a high V_t . The use of a high V_t is known as MTCMOS power gating [1].

For implementation of power gating circuits, there are many practical issues to be resolved. The outputs are floating in

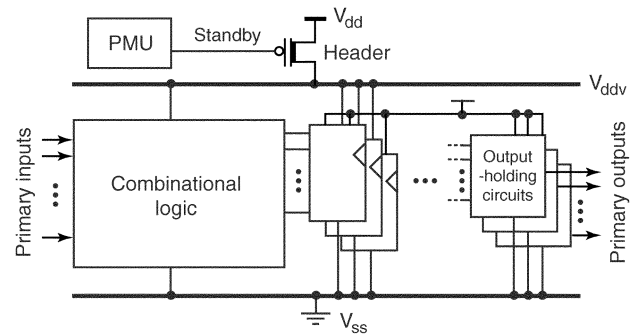


Fig. 1. Power gating circuits.

standby mode, which leads to a large short-circuit current in the blocks that are connected to the outputs, as well as logical errors in the outputs themselves. This can be alleviated by employing an interface circuit with the capability of preserving the logic during standby mode. As storage elements lose their states in standby mode, alternative elements, which are capable of state retention, must be used [1], [4], [5]. Sizing of the current switch is critical in terms of performance, area, and leakage current [6]. The large amount of current flowing into the current switch has an undesirable effect such as di/dt noise and long wakeup delay, thus large minimum sleep time [7]. These design issues specific to power gating, some of which call for tool support, make the application of power gating to semicustom designs difficult, especially for those based on standard-cell elements.

In addition to these issues, the physical design methodology needs to be tailored. A cell library specific to power gating needs to be designed to accommodate the requirement of additional power networks [1], [5]. The location of current switches and power-gating specific cells, such as state retention storage elements, is limited [8]–[10], and severely constrains the placement of logic cells. These problems are, in essence, due to the heterogeneous requirement on power networks. As an example, Fig. 1 shows that headers need V_{dd} and V_{ddv} ; logic cells are powered by V_{ddv} and V_{ss} ; state retention storage elements require all three power networks; output interface circuits need V_{dd} and V_{ss} . This can be avoided by embedding the current switch inside of each cell [2], however, a new cell library for this purpose needs to be developed.

In this brief, a semicustom design methodology for power gated circuits that enables the unmodified use of conventional standard-cell elements is proposed. In particular, a new power network architecture is proposed. A design method for a current switch is proposed in terms of substrate biasing and its placement; additionally, a new output-holding circuit is proposed, and this is compared to previous works. The proposed methodology is applied to ISCAS benchmark circuits and to a commercial Viterbi decoder with 0.18- μm CMOS technology.

Manuscript received July 18, 2006; revised November 27, 2006. This work was supported in part by ETRI and by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) under KRF-2005-003-D00247. This paper was recommended by Associate Editor J. Pineda de Gyvez.

The authors are with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea (e-mail: ppc750@kaist.ac.kr; youngsoo@ee.kaist.ac.kr).

Digital Object Identifier 10.1109/TCSII.2007.894414

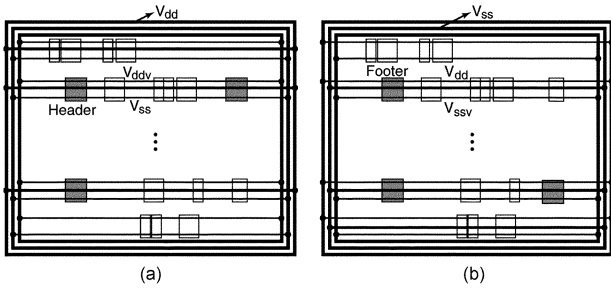


Fig. 2. Power networks for power gating circuits utilizing (a) headers and (b) footers.

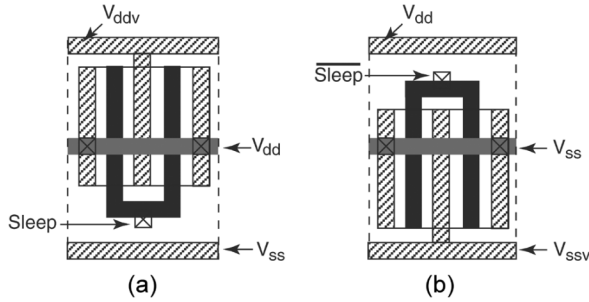


Fig. 3. Conceptual layout of current switch cells (a) header and (b) footer.

II. DESIGN OF POWER NETWORK AND CURRENT SWITCH

A. Power Network

Fig. 1 shows that additional power networks are needed for V_{ddv} as well as conventional networks for V_{dd} and V_{ss} . To meet this demand, the new power network topology shown in Fig. 2 is proposed. These networks consist of three power rings and corresponding power rails. When header switches are employed, a network of V_{dd} , V_{ss} , and V_{ddv} is constructed, where V_{dd} and V_{ss} networks are connected to chip-level power networks while V_{ddv} network is local. It is important to note that V_{ddv} and V_{ss} rails connect to the V_{dd} and V_{ss} terminals, respectively, of the cells implementing combinational logic, allowing unmodified conventional standard-cell logic elements to be used. Note also that V_{dd} rails have to reside in a higher metal layer to avoid any electrical connection to the logic cells, which they run across. As this higher metal layer may be reserved for signal routing, sharing the same layer with V_{dd} rails can increase wiring congestion. However, a selective use of V_{dd} rails (e.g., one rail per three cell rows) can alleviate the problem. This is discussed in Section IV in conjunction with the experimental results on the total wirelength. The networks using footer switches are shown in Fig. 2(b).

Fig. 3(a) shows the conceptual layout of a header cell. Its source and drain terminals are connected to V_{dd} and V_{ddv} , respectively, while its VSS terminal merely serves as a connecting medium for the cells on its left and right sides. It can be readily seen that the header cell, when placed as shown in Fig. 2(a), ensures the power gating structure of Fig. 1. Furthermore, its placement is not restricted as far as the V_{dd} rail is available to it; thus, it provides flexibility to the placement process. In the same context, the placement of power gating-specific cells (state retention storage elements and output interface circuits, which are addressed in Section III) is not restricted. The same advantages hold for the footer switch shown in Fig. 3(b).

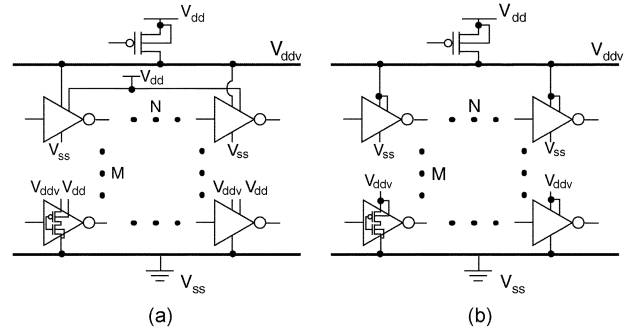


Fig. 4. Sizing of header switches with different substrate bias of logic cells.

B. Current Switch

1) *Substrate Biasing of Logic and Switch Sizing*: The seamless use of conventional standard cell elements implies that the proposed power gating structure is free from a body effect, as the unmodified standard cell elements have their sources and substrates tied together. As an example in Fig. 2(a), the source of pMOS devices, for instance of an inverter, is connected to V_{ddv} , where the n-well is also biased. As the delay of CMOS circuits increases with the body effect, the proposed power gating circuits are generally faster. This in turn implies that it is possible to use a smaller current switch for the same performance target.

Fig. 4 shows simulation circuits that aid with the understanding of the influence of body biasing of the logic cells on the size of the header switch. The logic block consists of M inverter chains, with each one having N inverters. In Fig. 4(a), the substrate of all pMOS devices are biased to V_{dd} (their sources connected to V_{ddv}), while it is biased to V_{ddv} in Fig. 4(b). This represents the proposed power gating structure. M and N are varied in order to change the charging patterns of the circuits, while the number of total inverters is held at 32.

When the delay penalty is set to 10% with M being equal to 1 (representing the case of the minimum charging current through the header switch), the size of the switch in Fig. 4(b) is approximately 3.4% smaller than that of Fig. 4(a) (V_{ddv} drops by 234 and 245 mV, respectively, with almost the same current of about 0.251 mA). The difference becomes smaller as M is increased, with a 1.3% difference when $M = 32$. The effect of the substrate biasing of logic gates on the footer switch is significant. With the same delay penalty of 10%, the size of the footer switch with the substrate of nMOS of the logic gates biased to V_{ssv} is smaller than that with the substrate of nMOS biased to V_{ss} in the range of 18% to 12%.

2) *Substrate Biasing and Layout of Switch*: While the substrate biasing of logic cells is implicit, as no standard cell layouts are modified, the substrate of the current switch can be biased either to its drain or to its source. This option implies a trade-off between area overhead and leakage saving. Fig. 5 shows two power gating structures with header switches. The header in the first architecture shares the n-well with logic gates, which enables a compact layout. However, during standby mode, the virtual V_{dd} does not drop completely, and the leakage in the circuit still remains. In contrast, the second architecture has an advantage in terms of standby mode leakage; the n-well of the header needs to be isolated from that of the logic gates, which increases the area overhead of the header switches.

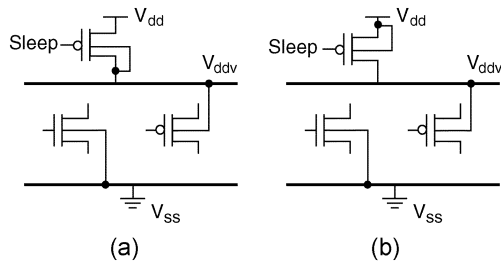


Fig. 5. Substrate biasing of a header switch.

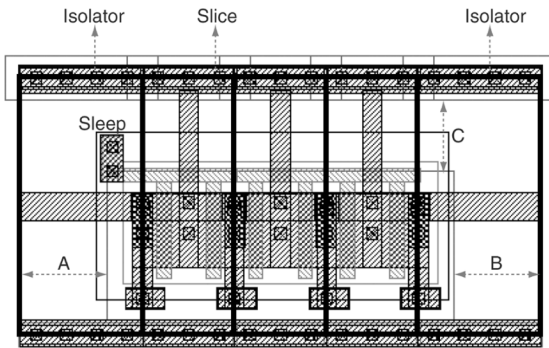


Fig. 6. Layout of a header switch with slices and isolators.

In order to evaluate these two options of substrate biasing, we take c5315, one of ISCAS benchmark circuits, which consists of 2335 gates after mapping it on to a commercial 180-nm, 1.8-V gate library. The header is sized assuming that 10% delay penalty can be tolerated. The virtual V_{dd} in the first type of substrate biasing Fig. 5(a) drops only to 1.36 V, while that in the second type drops to almost 0 V. The leakage in the first case is about 6090 nA; the leakage in the second case is only 3 nA. The cell layouts of two types of headers were created, and they were placed and routed together with the original circuit to see the area. In the first type of substrate biasing, the header takes 1.4% of total area; in the second type, the header (taking n-well isolation into account) takes 3.9%. At this point, the discussion is constrained to the second architecture for the remainder of this brief.

Current switches need to be physically distributed over the region where logic cells are also placed, such that the current requirements from the logic are satisfied. Thus, if current switches are used with their wells isolated, the area overhead can be significant. To cope with this problem, a current switch was built by combining two types of cells, termed a *slice* and an *isolator*. A slice is a unit current switch; when they are abutted together, they constitute a larger current switch. Isolators are placed at both ends of the slices so that there is guaranteed to be enough room between the switch and the logic cells for well isolation. Fig. 6 shows a header constructed by abutting three slices with two isolators. The spaces inside the isolators, denoted by A and B, guarantee the minimum spaces for the n-well of the slices and the n-wells of the logic cells, which are in different potentials. The space denoted by C provides well isolation between the slices and the logic cells placed on top of the slices with their orientation inverted.

Once the size (width) of a current switch has been determined [6] from a given performance requirement, the number of slices that need to be placed becomes apparent. In terms of a simple tally of area, the best means of placing slices is to abut them all

together, as this requires only two isolators. However, a single large current switch can block the placement of the logic cells. Furthermore, the power network (i.e., V_{ddv} or V_{ssv}) may experience a large IR drop if the logic cells are physically distant from the switch. On the other hand, if the slices are distributed such that they reduce the IR drop and avoid placement blockage, the area overhead increases, as two isolators are needed for each chunk of slices. This is termed a *slice block*.

To aid in the understanding of the area overhead for different styles of header placement, the ISCAS benchmark c7552, a combinational circuit with 3874 gates, is utilized. The number of slices is determined for different threshold voltages of headers with 0.18- μm CMOS technology, while the delay penalty is varied. For an example of a header switch with a high V_t and with 10% delay penalty, 117 slices are needed. The area overhead ranges from 2% (one slice block; thus, two isolators in total) to 7% (117 slice blocks with 234 isolators). If the delay penalty is 5%, making the performance requirement tighter, 213 slices are needed, and the area overhead becomes significant (3.6% to 12.5%). This suggests an interesting optimization problem that involves deciding upon the number of slice blocks and their placement [11], such that the power network is resilient to an IR drop and the area overhead remains within tolerance. This, however, is beyond the scope of this brief and is left for future studies.

III. POWER GATING SPECIFIC CELLS

When power gating circuits are in standby mode, outputs are floating. This leads to a large short circuit current in the blocks that are connected to the outputs. Thus, a special circuit is needed, termed an *output-holding circuit*, for each output, for the output to be held during standby mode.

Fig. 7(a) shows a circuit that can be used for power gating circuits with header switches.¹ During an active operation (i.e., sleep is de-asserted), M1 is always on, while either M2 or M3 is on depending on the output. A rising input is propagated through M1 and M2, which together behave as a transmission gate (note that M2 is initially on), thus the rising delay is small. A falling input is driven by M1 and M3, meaning that the delay is determined by the size ratio of M1 and M3. As M1 is already of a minimum size, in order to make M3 weaker and thus reduce the falling delay, the gate length of M3 is increased, with its gate of a minimum width.

During standby mode, assuming that the input is initially high and that the logic gates that drive this input are power-gated (sleep is asserted), hence the input gradually becomes lower. In this case, however, as both M1 and M2 are off, this change does not propagate to the output, while the output is held high by the pull-up pMOS device M3. Thus, the output is held high even when the logic gates are power-gated. The input of logic low is readily maintained, as it is not affected by the turned-off header switches. Note that, in our output-holding circuit, we assume that the virtual V_{dd} drops completely. When this is not true, for example when low V_t is used for a header, alternative circuits may be required. The proposed circuit has a larger amount of leakage when holding high, as M1, M2, and the pMOS of the

¹If a pull-down nMOS transistor is used instead of M3 and M1 and M2 are exchanged, an output-holding circuit is created, which is used for power gating circuits with footer switches.

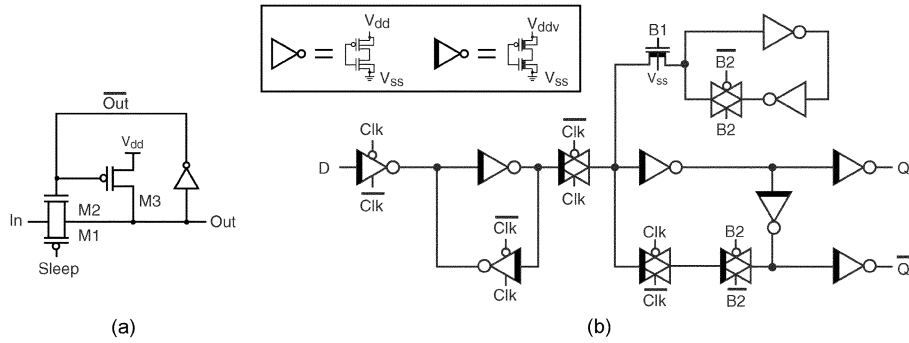


Fig. 7. (a) Output-holding circuit and (b) state-retention flip-flop.

TABLE I
COMPARISON OF OUTPUT INTERFACE CIRCUITS

Circuits	Area (μm^2)	Delay (ps)	Leakage (pA)
Proposed circuit	1.78	670	19
Leakage feedback gate	1.66	180	10900
Floating prevention circuit	2.09	660	73

inverter are off; as opposed to when holding low, when the leakage sources are M1, M3, and the nMOS of the inverter with M3 is less leaky.

The proposed output-holding circuit was compared to a leakage feedback gate [12] and floating prevention circuit [5]. As shown in Table I, the leakage feedback gate (with header only) has the largest amount of leakage due to the use of low V_t MOS devices; it is most efficient in terms of area (the area is approximated by the sum of $length \times width$ of the MOS transistors), again due to the use of low V_t . Note that the original leakage feedback gate [12] assumes both header and footer, which still has the largest leakage of 81 pA; its area becomes the largest ($3.20 \mu\text{m}^2$). Comparing the proposed circuit and the floating prevention circuit, it is clear that the former is more viable than the latter in terms of area and leakage, with a nearly identical delay (measured with the load of four inverters).

As all internal nodes float during standby mode, the data stored in storage elements, such as flip-flops and latches, are lost. The simplest approach to solve this problem is to isolate the storage elements from the current switches, i.e., directly connect them to the V_{dd} and ground. As V_{dd} or V_{ss} can be accessed from the cells (see Fig. 2), the layout of conventional storage elements, which connect to virtual rails if not modified, can be readily re-designed. However, the leakage current from the storage elements can be significant in this case.

Instead, the conventional storage elements are redesigned such that data is retained while most of the internal logic is power-gated. Fig. 7(b) shows the flip-flop with this state retention capability [13]. The cross-coupled inverters with the transmission gate provide state retention, while the remainder of the flip-flop is power-gated (i.e., low V_t inverters are connected to current switches), thus limiting the leakage current during standby mode. The B1 and B2 signals should be provided by the PMU [13]. The designed flip-flop can reduce leakage current by a factor of 300 and 50 when its state is low and high, respectively, at the cost of an area increase of 68%. The delay increase is negligible. In a similar way, a latch with state retention was also designed.

IV. DESIGN FLOW AND EVALUATION

Once we have a gate-level netlist after synthesizing Register Transfer Level (RTL) description, we repeat the timing analysis on the netlist while we change V_{dd} , to see how much voltage drop (from nominal V_{dd}) can be tolerated for given timing constraints. The drop, thus obtained, is assumed to be across a current switch. In order to obtain the average current flowing through the current switch, random patterns are applied and the netlist is then simulated. This combined with the voltage drop gives the size of current switch [6], which in turn renders the number of slices that need to be placed. An output-holding circuit is inserted at each primary output, and all of the storage elements in the netlist are replaced by state retention flip-flops or latches.

In the physical design stage, the conventional power/ground networks are initially generated. These networks, combined with the extra network for V_{dd} or V_{ss} , result in the power networks shown in Fig. 2. In the experiment of the power gating with header switches, M3 is utilized for the V_{dd} rails. The slice blocks are then placed in a regular fashion and are fixed in their locations. The placement of the slice blocks themselves need to be iterated with power network analysis [11], which is beyond the scope of this brief. After the placement of the logic cells, redundant V_{dd} rails, which are not connected to any headers, state retention storage elements, or output-holding circuits, are removed. The signal routing as well as the routing of the sleep signal follows. The transistor-level netlist is extracted from the layout, and is simulated with SPICE to estimate leakage current.

The experiments were performed on seven ISCAS benchmark circuits: four combinational and three sequential circuits. The results with industrial $0.18\text{-}\mu\text{m}$, 1.8-V CMOS technology are summarized in Table II. The low and high V_t are approximately 0.40 and 0.65 V respectively for nMOS (-0.42 and -0.60 V for pMOS). From the second to the fourth column are shown the characteristics of the original circuits. The remaining columns show the result after power gating. Header switches of a high V_t pMOS device are used in an isolated n-well. Output-holding circuits and parts of state retention flip-flop [refer to Fig. 7(b)] use high V_t as well. Metal layers are used up to M4 for routing and for a power network.

The area increase of the combinational circuits is due to header switches and output-holding circuits. The sizes of the header switches are dependent on the delay penalty, which is held constant to 10% for all circuits. Thus, the area overhead

TABLE II
EXPERIMENTAL RESULT ON ISCAS BENCHMARK CIRCUITS

Circuits	Original circuit			After power gating		
	Out-puts	SEs	Cells	Δ Area	Δ Wire-length	Leakage saving
c3540	22	0	1597	5.9%	9.0%	719×
c6288	32	0	1926	7.8%	-5.5%	1071×
c5315	109	0	2335	14.6%	0.7%	234×
c7552	100	0	3874	8.0%	4.0%	303×
s820	19	5	510	13.1%	14.3%	1215×
s1423	5	74	861	20.7%	28.3%	545×
s9234	20	145	2101	18.0%	12.7%	542×

is larger for the circuits with more outputs. State retention storage elements are another component of area overhead for the sequential circuits.

There are three main sources for the wirelength increase (sixth column): the control signals for flip-flops (B1 and B2); the control signals for current switches and output-holding circuits (sleep); and the increased signal wires due to routing congestion. The first source, which takes the largest proportion, only applies to the sequential circuits. Correspondingly, the sequential circuits have a greater increase in wirelength. The increase of signal wires may affect dynamic power and circuit delay. For s1423, whose total wirelength increases the most, dynamic power and circuit delay increases about 1.5% and 0.8%, respectively.

The last column in Table II shows the leakage saving factor compared to non-power-gated circuit of mixed V_t .² During standby mode, the leakage components are header switches, output-holding circuits, and state retention storage elements, with the header switches being a minor component of the total leakage current.

To test transition behavior of our power gating circuits, we take c5315 as an example. It takes 795 ns and 2 ns for sleep and wakeup. The total energy for going sleep and waking up is 9.93 nJ. This gives us 4.6 μ s of minimum idle time, below which power gating does not yield power saving.

V. CASE STUDY: VITERBI DECODER

In order to validate the proposed power gating methodology, a Viterbi decoder is used as a test vehicle. The Viterbi decoder is a core module in a mobile-station baseband modem, whose standby power consumption is of critical importance. It is widely used to decode convolutional codes, and is very popular to correct errors from a communication channel.

The design used in this experiment is designed in VHDL, and follows the design flow in the previous section with the physical design done in flat. The decoder works at a maximum data-rate of 500 kb/s at 100 MHz. The decoder consists of 6475 cells, where 1549 are storage elements.

It is important to note that the decoder is dominated by storage elements, although it does not have a large number of outputs, which explains the relatively large increase in area (28.4%) and wirelength (22.9%). In order to see the impact of additional V_{dd} rails (on M3), which we have in all 113 circuit rows due to large number of storage elements that are spread out after placement,

²We implemented a gate-level mixed V_t algorithm [14] to convert original all low V_t circuits to mixed V_t one.

we remove them and re-route the whole design. The total wire-length (not including V_{dd} rails) remains almost constant. This is because V_{dd} rail occupies two tracks out of nine tracks of each circuit row and the remaining seven tracks are more than enough for routing on M3 layer.

VI. CONCLUSION

The application of power gating circuits to semicustom designs based on standard-cell elements is limited due to the requirement of developing standard cells that are tailored for power gating or the requirement of customizing physical design methodologies. A design method of power network is proposed in this brief that enables use of conventional standard-cell elements without customization. The approach is free from the body effect, and the impact of this is investigated in terms of the switch size. A method of current switch design is discussed focusing on the way the layout is constructed. A new output-holding circuit is proposed, and is shown to be superior to those in previous works. The proposed design methodology was applied to ISCAS benchmark circuits, and also to a Viterbi decoder with industrial 0.18- μ m CMOS technology.

REFERENCES

- [1] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "A 1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS," *IEEE J. Solid-State Circuits*, vol. 30, no. 8, pp. 847–854, Aug. 1995.
- [2] K. Usami, N. Kawabe, M. Koizumi, K. Seta, and T. Furusawa, "Automated selective multi-threshold design for ultra-low standby applications," in *Proc. Int. Symp. Low Power Electron. Design*, Aug. 2002, pp. 202–206.
- [3] T. Kitahara, N. Kawabe, F. Minami, K. Seta, and T. Furusawa, "Area-efficient selective multi-threshold CMOS design methodology for standby leakage power reduction," in *Proc. Design, Automat. and Test in Europe Conf. Exhib.*, Mar. 2005, pp. 646–647.
- [4] V. Zyuban and S. V. Kosonocky, "Low power integrated scan-retention mechanism," in *Proc. Int. Symp. Low Power Electron. Design*, Aug. 2002, pp. 98–102.
- [5] H.-S. Won, K.-S. Kim, K.-O. Jeong, K.-T. Park, K.-M. Choi, and J.-T. Kong, "An MTCMOS design methodology and its application to mobile computing," in *Proc. Int. Symp. Low Power Electron. Design*, Aug. 2003, pp. 110–115.
- [6] S. Mutoh, S. Shigematsu, Y. Gotoh, and S. Konaka, "Design method of MTCMOS power switch for low-voltage high-speed LSIs," in *Proc. Asia South Pacific Design Automat. Conf.*, Jan. 1999, pp. 113–116.
- [7] S. Henzler *et al.*, "Sleep transistor circuits for fine-grained power switch-off with short power-down times," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2005, pp. 302–303.
- [8] S. V. Kosonocky, M. Immediato, P. Cottrell, and T. Hook, "Enhanced multi-threshold (MTCMOS) circuits using variable well bias," in *Proc. Int. Symp. Low Power Electron. Design*, Aug. 2001, pp. 165–169.
- [9] P. Babighian, L. Benini, A. Macii, and E. Macii, "Post-layout leakage power minimization based on distributed sleep transistor insertion," in *Proc. Int. Symp. Low Power Electron. Design*, Aug. 2004, pp. 138–143.
- [10] P. Royannez *et al.*, "90-nm low leakage SoC design techniques for wireless applications," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2006, pp. 138–139.
- [11] J. N. Kozhaya and L. A. Bakir, "An electrically robust method for placing power gating switches in voltage islands," in *Proc. Custom Integr. Circuits Conf.*, Oct. 2004, pp. 321–324.
- [12] J. Kao and A. Chandrakasan, "MTCMOS sequential circuits," in *Proc. Eur. Solid-State Circuits Conf.*, Sep. 2001, pp. 317–320.
- [13] S. Shigematsu, S. Mutoh, Y. Matsuya, Y. Tanabe, and J. Yamada, "A 1-V high-speed MTCMOS circuit scheme for power-down application circuits," *IEEE J. Solid-State Circuits*, vol. 32, no. 6, pp. 861–869, Jun. 1997.
- [14] M. Ketkar and S. S. Sapatnekar, "Standby power optimization via transistor sizing and dual threshold voltage assignment," in *Proc. Int. Conf. on Computer Aided Design*, Nov. 2002, pp. 375–378.